AD_____

Award Number: DAMD17-03-1-0212

TITLE: Computerized Analysis and Detection of Missed Cancer in Screening Mammogram

PRINCIPAL INVESTIGATOR: Lihua Li, Ph.D.

CONTRACTING ORGANIZATION: H. Lee Moffitt Cancer Center and Research Institute
Tampa, FL 33612

REPORT DATE: April 2004

TYPE OF REPORT: Annual Summary

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	April 2004	Annual Summary (24 Mar 2003 - 23 Mar 2004)

4. TITLE AND SUBTITLE

Computerized Analysis and Detection of Missed Cancer in Screening Mammogram

5. FUNDING NUMBERS

DAMD17-03-1-0212

6. AUTHOR(S)

Lihua Li, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

H. Lee Moffitt Cancer Center and
Research Institute
Tampa, FL 33612

E-Mail: lilh@moffitt.usf.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 Words)

This project is to explore an innovative CAD strategy for improving early detection of breast cancer in screening mammograms by focusing on computerized analysis and detection of cancers missed by radiologists. The research scope in past year is on database generation and analysis of missed cancers. Several major progresses have been made including (1) By reviewing more than 1334 cases, a total of 83 missed cancer cases were collected which were used to generate three different datasets including mammograms with missed cancer, mammograms with screening-detected cancer and normal mammograms. (2) Regions-of-interest (ROIs) containing a detected or a missed cancer were extracted, and a ground truth was generated by an experienced radiologist for feature extraction and analysis purpose. (3) With the datasets and the ground truth, a variety of computerized features were extracted and analyzed to explore the difference of detected and missed cancer cases. A set of tests was applied to the extracted features individually from which the significant features distinguishing the missed cancer from detected ones could be identified and applied potentially to the CAD design in next steps.

14. SUBJECT TERMS

Breast Cancer, Mass, Detection, CAD, Mammography, Robust, Adaptive, Segmentation, Classification

15. NUMBER OF PAGES

16

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

INTRODUCTION

This project is to explore an innovative CAD strategy for improving early detection of breast cancer in screening mammograms by focusing on computerized analysis and detection of cancers missed by radiologists. As listed in the Statement of Work, the research scope in the first year of project is to generate databases and analyze the missed cancers.

BODY

Objective 1: ***to generate databases for missed cancer analysis and detection.***

Accomplishments:

1. Data Collection Criteria and Procedure

a. The criteria for inclusion in this study were as follows:
   1. Mass must be visible on mammogram
   2. Mass must be proven by biopsy to be malignant
   3. Mass must be seen in retrospect on a prior mammogram when reviewed by a radiologist

b. Procedure used for case selection:
   1. Lists of patients from both the screening and diagnostic centers were obtained
   2. Each patient's chart was reviewed to select for masses that were visible mammographically, all others were excluded
   3. The selected cases were reviewed for malignant pathology outcome, all others were excluded
   4. Films were requested from the diagnostic center for those cases with malignant masses
   5. Films from the screening center had to be obtained manually due to lack of manpower
   6. Films were reviewed to ascertain whether the exam and prior mammograms were available. Only those with prior mammograms were selected.
   7. Selected mammograms were reviewed by a radiologist to determine a) if the mass was visible retrospectively on the prior exam and b) the reason it was not detected on the prior exam
   8. The radiologist indicated the location and outlined the contour of the lesion on both exams and the Breast Imaging Reporting And Data System (BI-RADS) descriptors
   9. Ground truth files (hard copy) were generated based on the radiologists outlines
   10. The films were then digitized manually on a Kodak (LUMISYS) LS85 digitizer at a resolution of 50μm and 12 bits in grey scale.

2. Sources and number of cases reviewed: (as of March 23, 2004)

Query of patient databases	770
Staging database	93
Teaching files archive	148
Breast conference patients	100
Log of invasive procedures	160
Research archives	63
Total number of cases reviewed	**1,334**

3. Reasons for exclusion of cases from the original 1,334 patients reviewed:

Duplication of names among lists
Lesion was something other than a mass

Lesion was a benign mass
No pathology available
No information available for this patient/exam
No follow up for this patient
Films were unavailable or incomplete
Mass was not visible on prior mammogram (interval cancer)

a. Analysis of the 770 names from patient database queries:

Reason	**Number excluded**
Duplication of names among lists	49
Lesion was something other than a mass	337
Lesion was a benign mass	111
No information available	51
No follow up available	56

This leaves a balance of 166 potential cases, of which:

Films were unavailable or incomplete	100
Mass not visible on prior exam	16
Miscellaneous exclusions	21
Usable cases	**29**

b. Analysis of the 93 names from the staging database:

Reason	**Number excluded**
Duplication of names among lists	1
Lesion was something other than a mass	39
No information available	9

This leaves a balance of 44 potential cases, of which:

Films were unavailable or incomplete	42
Usable cases	**2**

c. Analysis of the 148 names from teaching files:

Reason	**Number excluded**
Duplication of names among lists	20
Lesion was something other than a mass	58
Lesion was a benign mass	12
No information available	13
No pathology available	1

This leaves a balance of 44 potential cases, of which:

Films were unavailable or incomplete	32
Mass not visible on prior exam	5
Usable cases	**7**

d. Analysis of the 100 names from breast conference lists:

Reason	**Number excluded**
Duplication of names among lists	8
Lesion was something other than a mass	34
Lesion was a benign mass	1

No information available 12

This leaves a balance of 45 potential cases, of which:

Films were unavailable or incomplete	29
Mass not visible on prior exam	4
Usable cases	**12**

e. Analysis of the 160 names from invasive procedures log:

Reason	**Number excluded**
Duplication of names among lists	4
Lesion was something other than a mass	71
Lesion was a benign mass	4
No information available	20

This leaves a balance of 61 potential cases, of which:

Films were unavailable or incomplete	34
Mass not visible on prior exam	5
Usable cases	**22**

f. Analysis of the 63 names from research archives:

Reason	**Number excluded**
Duplication of names among lists	2
Lesion was something other than a mass	22
Lesion was a benign mass	5
No pathology available	9

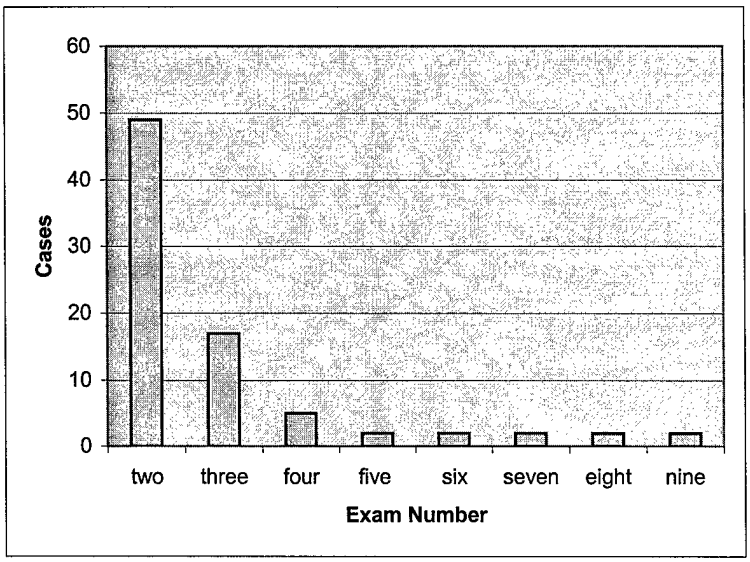
This leaves a balance of 25 potential cases, of which:

Mass not visible on prior exam	11
Usable cases	**14**

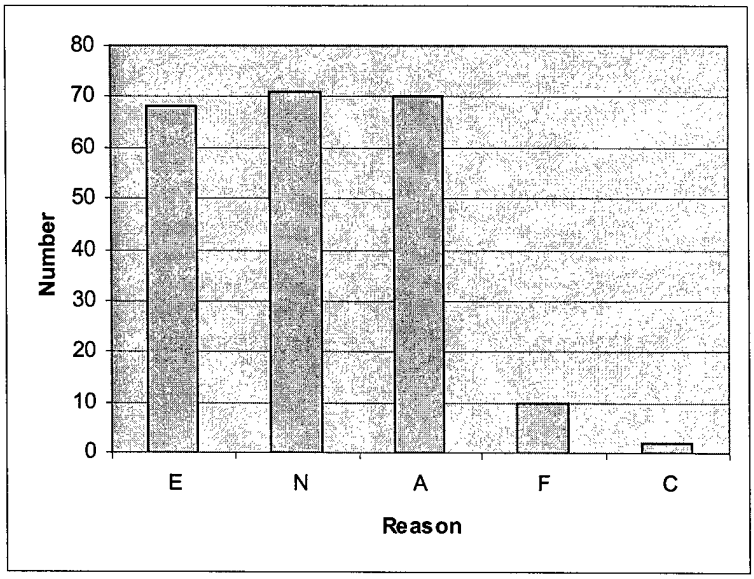
Summary: As of March 23, 2004, a total of 86 out of 1334 cases were collected as missed cancer cases for study. It is projected that there will be another 20 cases be collected before the end of May 2004, so that the total number of missed cancer cases will be more than 100.

4. Characteristic analysis of the database

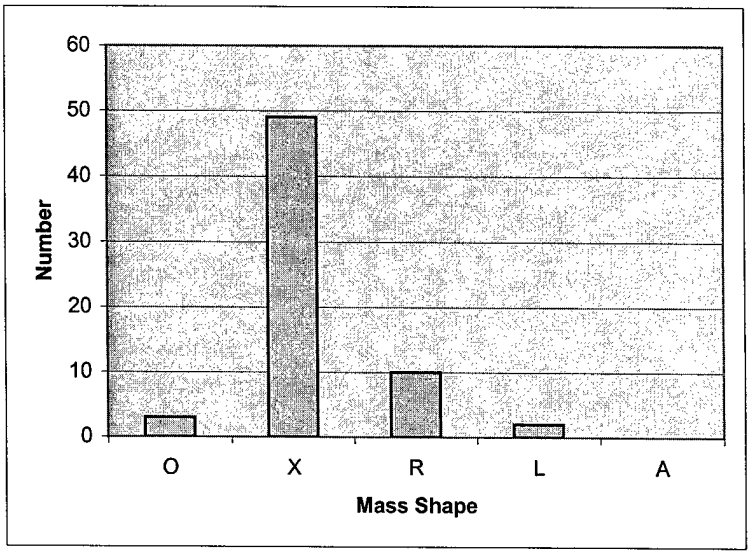
The characteristics of database was analyzed by following descriptions: (a) Case distribution in terms of exam numbers, (b) Case distribution in terms of cancer missed reasons (per view and stage), (c) Case distribution in terms of mass shape, (d) Case distribution in terms of mass margin, (e) Case distribution in terms of Mass density. The histograms are shown in Figure 1.


60
50
40
30
20
10
0
Cases
two
three
four
five
six
seven
eight
nine
Exam Number

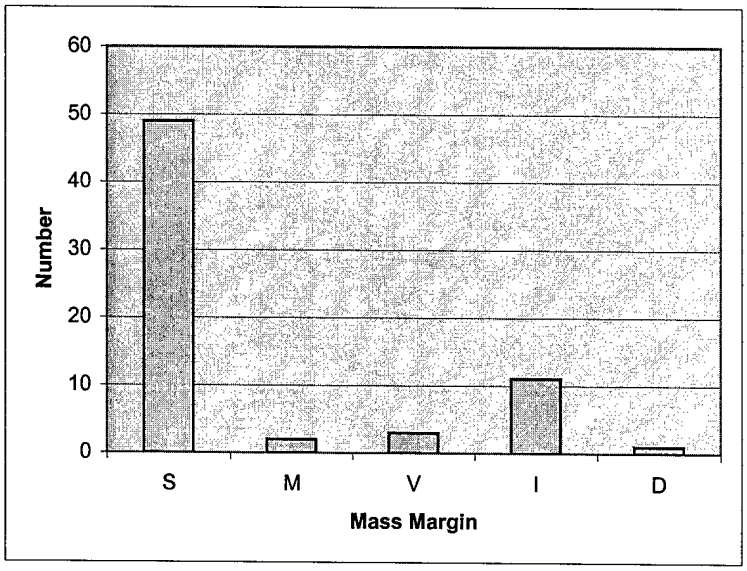
(a)


80
70
60
50
40
30
20
10
0
Number
E
N
A
F
C
Reason

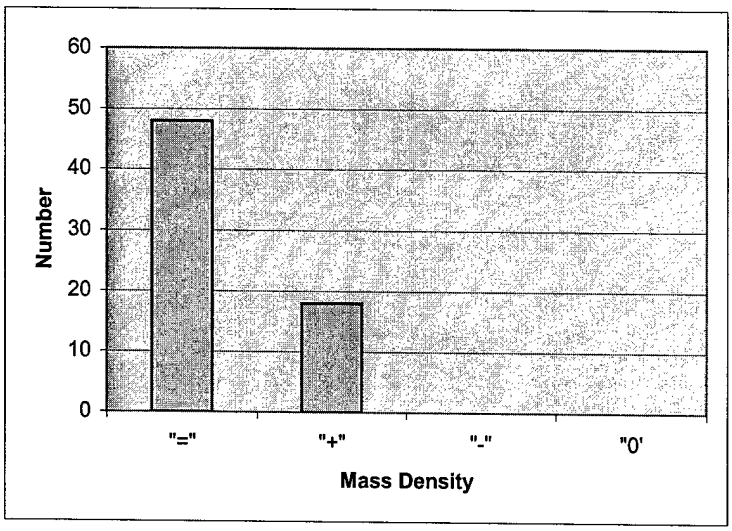
(b)


60
50
40
30
20
10
0
Number
O
X
R
L
A
Mass Shape

(c)



(d)



(e)

Figure 1. Case distribution in terms of (a) exam numbers, (b) missed reasons (E-interpretation error, N-not significant evidence, A-absent/no sign, F-not in field of view, C-contrast problem), (c) mass shape (O-oval, X-irregular, R-round, L-lobulated, A-architectural distortion), (d) mass margin (S-spiculated, M-microlobulated, V-obscured, I-indistinct ill defined, D-circumscribed well defined/sharply defined), (e) Mass density (=: equal/isodense, +: high, -: low, 0: fat containing/radiolucent).

Objective 2: to analyze the computerized features of missed cancers (false negatives) versus detected ones (true positives)

Accomplishments:

1. Data preprocessing

There are totally 86 cases of series mammograms in the database now. Due to the

difficulty and time consuming of data collection as described above and the research timeline limitation, some preprocessing and missed cancer analysis work had to be taken in parallel with data collection. In this feature analysis study, 73 cases were processed. More and/or complete analysis will be followed. The preprocessing work for data analysis includes image format transformation (from Digital Imaging and Communications in Medicine (DICOM) format to Sun TAAC Image File Format (VFF)), image re-sampling for mass feature extraction purpose (from 50 μm to 200 μm).

2. Mass feature analysis: missed vs. detected

(1) ***ROI generation:*** Based on the mass location (center) indicated by radiologist, two sets of regions-of-interest (ROIs) are created with 256×256 pixels in size. One contains a detected mass in each ROI, the second set consists of ROIs with missed masses.

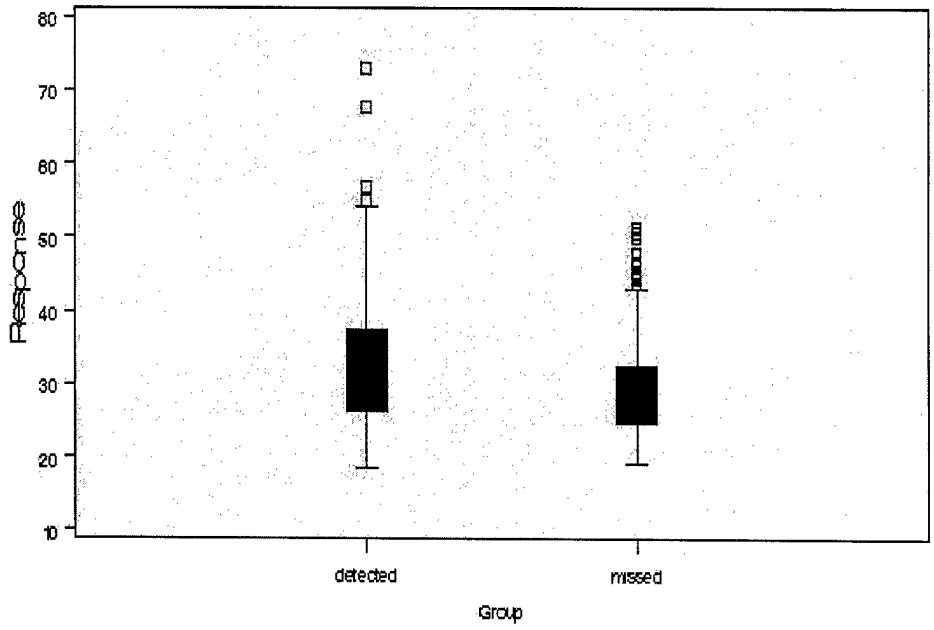
(2) ***Mass segmentation:*** Based on the ground truth (mass contour) generated by an experienced mammographer, a manual segmentation of the mass was taken by following the outline interactively with a tool we developed under Interactive Data Language (IDL) environment.

(3) ***Feature calculation:*** Following features are designed and calculated on both detected and missed masses using the original ROI image and the segmented image [1]:
Gray-level features: Intensity Mean, Intensity Variance, Intensity difference between mass area and surrounding background area;
Morphological features: Size, Circularity, Compactness, Roughness, Fluctuation, FWHM (Full-Width Half-Maximum), Radial gradient;
Texture features: Generalized Co-occurrence Matrix (GCM) based features (Energy, Difference moment, Inverse difference moment, Correlation), Laws features.

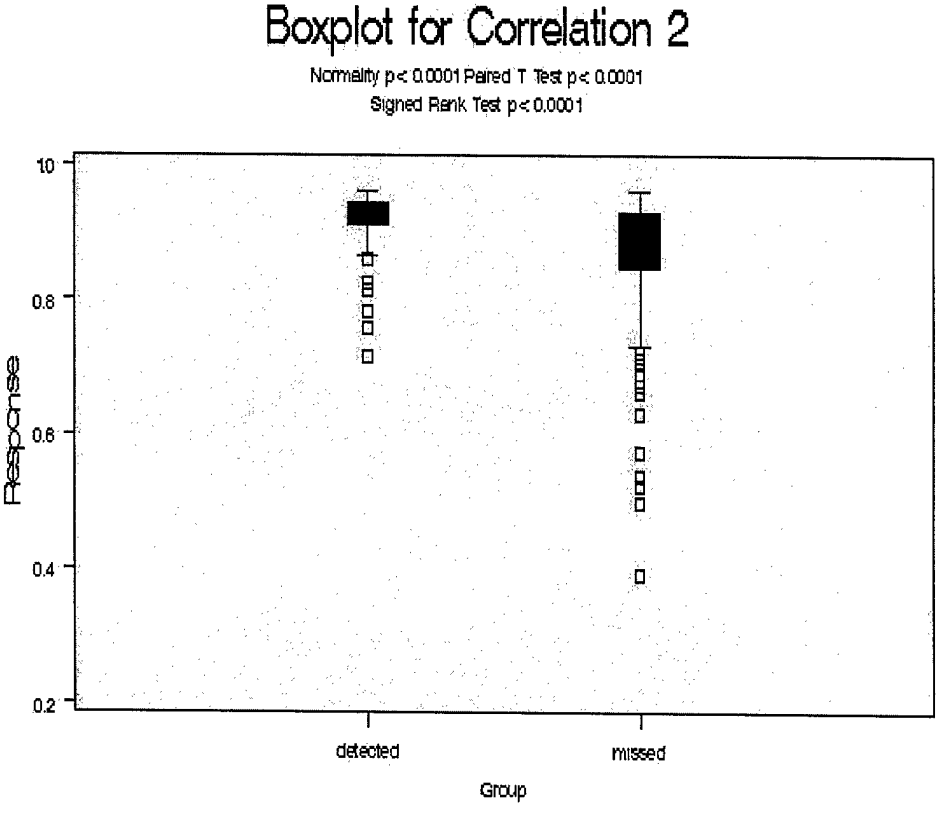
(4) ***Statistical analysis:*** To explore the difference of detected and missed cancer features, a set of tests was applied to the extracted features individually. Listed in Table 1 are the *p*-values of three tests including normality test, paired t-test, and signed rank test for each feature [2]. In order to explore the potential effect of mammography exam view on interpretation and the difference of missed cancer features on different views, in addition to the Craniocaudal (CC) and Mediolateral Oblique (MLO) combined test, statistical tests on CC view only and MLO view only were also taken. Following is the interpretation of test results:

- If normality p-value is less than 0.05, we say the difference between miss and detection of certain feature is not normally distributed.
- If the difference between miss and detection of certain feature is normally distributed, we use paired t-test. If t-test P-value is less than 0.05, we have evidence to reject null hypothesis that the mean of difference is zero at significant level 0.05. (significantly different)
- If the difference between miss and detection of certain variable is not normally distributed, we use signed rank test. If signed rank test P-value is less than 0.05, we have evidence to reject null hypothesis that the mean of difference is zero at significant level 0.05. (significantly different)
- From the table, the most significantly changed features are size, intensity variance, intensity difference, compactness, correlations, difference entropy, and inverse difference moments.

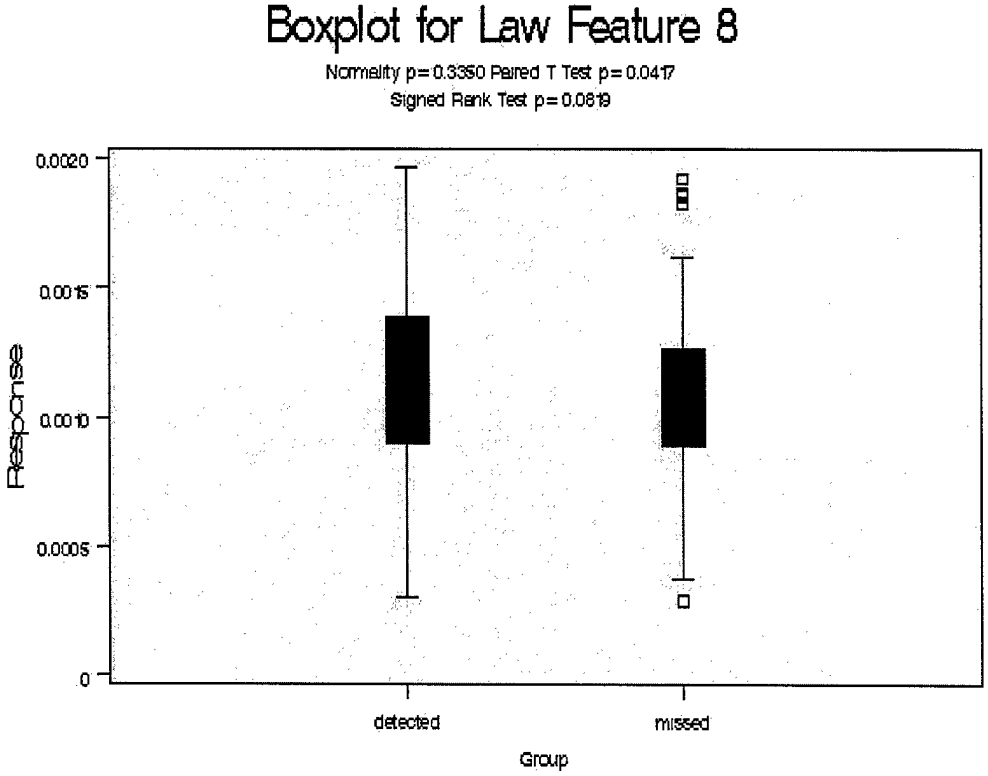
For illustrative purpose, box-plots of four features are shown in Figure 2. It is observed that the features of Compactness and Correlation 2 (at 45 degree) have a significant difference between the detected and missed masses, while there are not statistical difference in terms of Laws Feature 8 and intensity Mean.


Boxplot for Compactness
Normality p=0.0002 Paired T Test p= 0.0002
Signed Rank Test p=0.0008
Response
80
70
60
50
40
30
20
10
detected
missed
Group

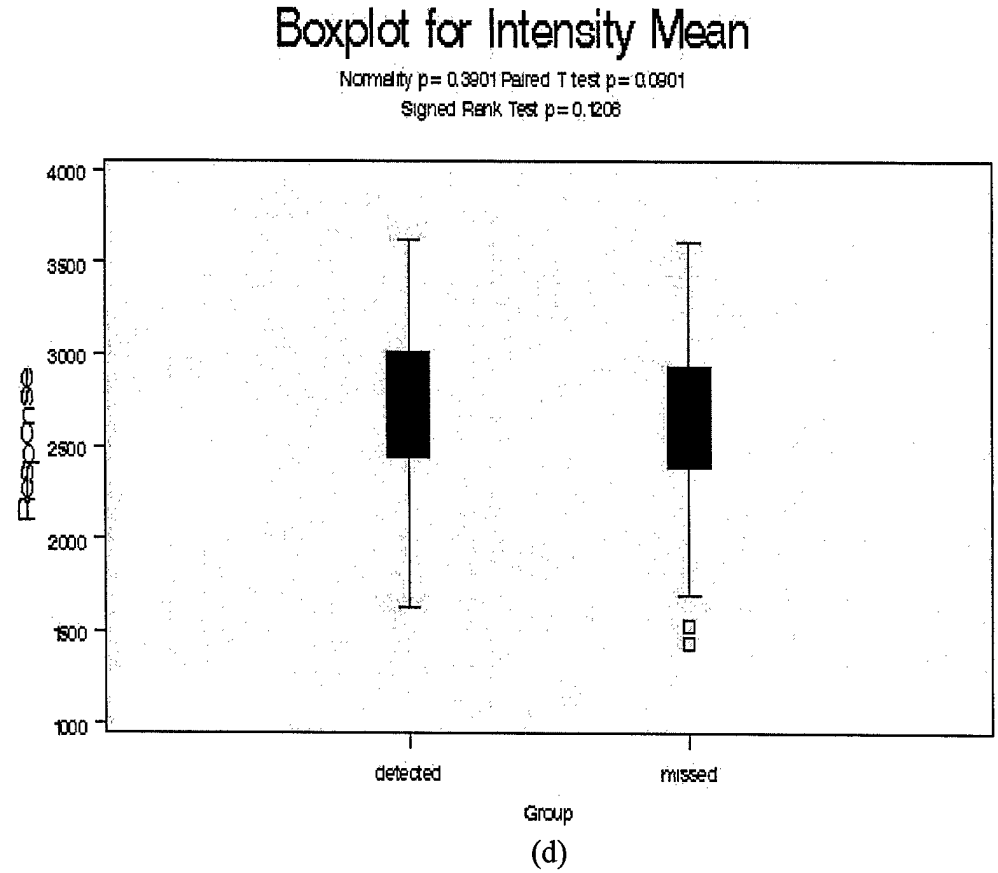
(a)


Boxplot for Correlation 2
Normality p< 0.0001 Paired T Test p< 0.0001
Signed Rank Test p< 0.0001
Response
1.0
0.8
0.6
0.4
0.2
detected
missed
Group

(b)



(c)



(d)

Figure 2. Box-plots for the illustration of statistical tests of the difference of four computerized features between missed and detected cancers.

3. Breast density analysis

(1) The breast area in a mammogram is segmented from the surrounding background. The chest wall is removed by manual segmentation. Based on the characteristic features of the gray level histogram of breasts at different intensity level, a gray level threshold value for each image is determined by interactive method to segment the dense area from the breast. Four classes can be classified according to a gray level histogram of the breast area. A typical Class I is almost entirely fat, it has a single narrow peak on the histogram. Class II has scattered fibroglandular densities. It has two peaks. The smaller peak is on the right of the bigger one. Class III is heterogeneously dense. It has two peaks, but the smaller peak is on the left of the bigger one. Class IV is extremely dense, which has a single dominant peak on the histogram, but it is wider compared with the peak in the Class I histogram.

(2) The area of segmented dense tissue as a percentage of the breast area is then calculated as the index of breast density.

(3) A preliminary study was taken to analyze the breast density feature of missed cancer cases versus detected cases. The p-values of statistical test are listed in Table 1.

4. Temporal Analysis

Temporal analysis was taken to explore the difference of characteristics between the changes of features among normal region, missed cancer region and detected cancer region. Following features of each ROI are calculated [1]: (1) Intensity Mean, (2) Intensity Variance, (3) Energy, (4) Difference Moment, (5) Inverse Difference Moment, (6) Correlation, and (7) 14 Laws features. Listed in Table 1 are the *p*-values of three tests including normality test, paired t-test, and signed rank test for each feature [2].

Table 1. P-Value Table: Missed vs. Detected

FEATURE NAME	VIEW	NORMALITY	PAIRED T TEST	SIGNED RANK TEST
Size	CC & MLO	<0.0001	<0.0001	<0.0001
	CC	0.0017	<0.0001	<0.0001
	MLO	<0.0001	<0.0001	<0.0001
Intensity Mean	CC & MLO	0.3901	0.0901	0.1206
	CC	0.3430	0.1864	0.2675
	MLO	0.9198	0.2961	0.3102
Intensity Variance	CC & MLO	<0.0001	<0.0001	<0.0001
	CC	0.9714	<0.0001	<0.0001
	MLO	<0.0001	<0.0001	<0.0001
Intensity Difference	CC & MLO	0.0020	<0.0001	<0.0001
	CC	0.0039	<0.0001	<0.0001
	MLO	0.2125	<0.0001	<0.0001
Circularity	CC & MLO	0.0058	0.2910	0.3514
	CC	0.2054	0.8544	0.9941
	MLO	0.0035	0.1815	0.1485
Compactness	CC & MLO	0.0002	0.0002	0.0006
	CC	0.0033	0.0026	0.0046
	MLO	0.0056	0.0239	0.0435
Roughness	CC & MLO	0.9990	0.7341	0.7418
	CC	0.8514	0.8370	0.7942
	MLO	0.9171	0.7785	0.8501

Fluctuation	CC & MLO	0.0305	0.3971	0.2200
	CC	0.0662	0.5970	0.3196
	MLO	0.0376	0.5091	0.5457
FWHM	CC & MLO	0.1922	0.8510	0.9160
	CC	0.3860	0.4120	0.3616
	MLO	0.1507	0.2451	0.4618
Radial Gradient	CC & MLO	0.0953	0.5127	0.3446
	CC	0.4060	0.2434	0.2047
	MLO	0.3737	0.8030	0.9189
Energy 1 (0°)	CC & MLO	<0.0001	0.3936	0.9370
	CC	<0.0001	0.5053	0.8580
	MLO	0.0004	0.5975	0.9652
Energy 2 (45°)	CC & MLO	<0.0001	0.6619	0.5762
	CC	<0.0001	0.6952	0.6120
	MLO	0.0002	0.8280	0.7991
Energy 3 (90°)	CC & MLO	<0.0001	0.4716	0.7709
	CC	<0.0001	0.5435	0.7656
	MLO	0.0001	0.6921	0.9247
Energy 4 (135°)	CC & MLO	<0.0001	0.6684	0.5407
	CC	<0.0001	0.6988	0.6015
	MLO	0.0001	0.8333	0.7712
Difference Moment 1 (0°)	CC & MLO	<0.0001	0.3298	0.0118
	CC	0.0048	0.3024	0.0721
	MLO	<0.0001	0.6863	0.0721
Difference Moment 2 (45°)	CC & MLO	<0.0001	0.6844	0.0518
	CC	0.0141	0.5397	0.2302
	MLO	<0.0001	0.9612	0.1159
Difference Moment 3 (90°)	CC & MLO	<0.0001	0.3230	0.0197
	CC	0.0028	0.2845	0.0525
	MLO	0.0010	0.6655	0.1706
Difference Moment 4 (135°)	CC & MLO	<0.0001	0.5049	0.0151
	CC	0.0002	0.4790	0.0733
	MLO	<0.0001	0.7580	0.1075
Inverse Difference Moment 1 (0°)	CC & MLO	0.5219	0.0006	0.0002
	CC	0.9513	0.0289	0.0232
	MLO	0.4463	0.0076	0.0024
Inverse Difference Moment 2 (45°)	CC & MLO	0.3035	0.0038	0.0010
	CC	0.9965	0.0601	0.0516
	MLO	0.1456	0.0264	0.0062
Inverse Difference Moment 3 (90°)	CC & MLO	0.0132	0.0019	0.0002
	CC	0.1402	0.0451	0.0151
	MLO	0.1016	0.0168	0.0040
Inverse Difference Moment 4 (135°)	CC & MLO	0.0402	0.0029	0.0004
	CC	0.0916	0.0490	0.0154
	MLO	0.2635	0.0272	0.0135
Correlation 1 (0°)	CC & MLO	<0.0001	<0.0001	<0.0001
	CC	<0.0001	0.0134	<0.0001
	MLO	<0.0001	<0.0001	<0.0001
Correlation 2 (45°)	CC & MLO	<0.0001	<0.0001	<0.0001
	CC	<0.0001	0.0006	<0.0001
	MLO	<0.0001	<0.0001	<0.0001
Correlation 3 (90°)	CC & MLO	<0.0001	<0.0001	<0.0001
	CC	<0.0001	0.0152	<0.0001
	MLO	<0.0001	<0.0001	<0.0001

Correlation 4 (135°)	CC & MLO	<0.0001	<0.0001	<0.0001
	CC	<0.0001	0.0033	<0.0001
	MLO	<0.0001	<0.0001	<0.0001
Laws Feature 1	CC & MLO	<0.0001	0.0337	0.0373
	CC	<0.0001	0.0970	0.0506
	MLO	0.4194	0.1912	0.3280
Laws Feature 2	CC & MLO	<0.0001	0.0866	0.0167
	CC	<0.0001	0.1364	0.0575
	MLO	0.0001	0.4029	0.1571
Laws Feature 3	CC & MLO	<0.0001	0.0856	0.0146
	CC	<0.0001	0.1356	0.0488
	MLO	<0.0001	0.3971	0.1485
Laws Feature 4	CC & MLO	<0.0001	0.0574	0.0484
	CC	<0.0001	0.0973	0.1425
	MLO	0.0712	0.3605	0.2218
Laws Feature 5	CC & MLO	0.5129	0.0841	0.0872
	CC	0.4619	0.2403	0.1838
	MLO	0.5717	0.2095	0.2963
Laws Feature 6	CC & MLO	0.0088	0.0346	0.0466
	CC	0.0028	0.1446	0.1383
	MLO	0.4038	0.1275	0.2081
Laws Feature 7	CC & MLO	0.0015	0.0275	0.0399
	CC	0.0010	0.1419	0.1692
	MLO	0.3080	0.0976	0.1464
Laws Feature 8	CC & MLO	0.3350	0.0417	0.0819
	CC	0.2689	0.1936	0.2515
	MLO	0.4877	0.1144	0.1899
Laws Feature 9	CC & MLO	<0.0001	0.0245	0.0299
	CC	<0.0001	0.1294	0.1404
	MLO	0.3082	0.0866	0.1195
Laws Feature 10	CC & MLO	<0.0001	0.0290	0.0509
	CC	<0.0001	0.1487	0.1941
	MLO	0.2991	0.0892	0.1527
Laws Feature 11	CC & MLO	0.0623	0.0539	0.1032
	CC	0.0550	0.2385	0.3196
	MLO	0.4846	0.1169	0.1729
Laws Feature 12	CC & MLO	<0.0001	0.0398	0.0862
	CC	<0.0001	0.1875	0.2911
	MLO	0.2861	0.0989	0.1777
Laws Feature 13	CC & MLO	0.1695	0.0630	0.0976
	CC	0.1234	0.2750	0.3159
	MLO	0.6673	0.1186	0.1660
Laws Feature 14	CC & MLO	0.6084	0.0839	0.0800
	CC	0.5726	0.3567	0.2842
	MLO	0.7555	0.1242	0.1108
Density	CC & MLO	0.0085	0.0230	0.3594
	CC	0.0413	0.5366	0.8522
	MLO	0.0946	0.0073	0.0199

Table 2 Temporal Comparison P-value

FEATURE NAME	Normality	Paired T-Test	Signed Rank Test
Intensity Mean	0.8584	0.0099	0.0069
Intensity Variance	0.1426	0.4962	0.3167
Energy 1 (0°)	0.9759	0.9445	0.8176
Energy 2 (45°)	0.9510	0.9592	0.8332
Energy 3 (90°)	0.9791	0.9562	0.8176
Energy 4 (135°)	0.9808	0.9378	0.8020
Difference Moment 1 (0°)	0.9001	0.4837	0.5001
Difference Moment 2 (45°)	0.3719	0.6939	0.6806
Difference Moment 3 (90°)	0.9847	0.3220	0.3799
Difference Moment 4 (135°)	<0.0001	0.3010	0.6513
Inverse Difference Moment 1 (0°)	0.9352	0.5495	0.6083
Inverse Difference Moment 2 (45°)	0.8829	0.8537	0.9441
Inverse Difference Moment 3 (90°)	0.8287	0.4730	0.4622
Inverse Difference Moment 4 (135°)	0.7900	0.4166	0.4378
Correlation 1 (0°)	<0.0001	0.2298	0.1328
Correlation 2 (45°)	<0.0001	0.2983	0.1274
Correlation 3 (90°)	0.0051	0.3962	0.2050
Correlation 4 (135°)	<0.0001	0.1911	0.1383
Laws Feature 1	<0.0001	0.3688	0.2075
Laws Feature 2	0.0107	0.0557	0.0152
Laws Feature 3	0.0007	0.1023	0.0196
Laws Feature 4	0.0443	0.0350	0.0140
Laws Feature 5	<0.0001	0.7859	0.0886
Laws Feature 6	<0.0001	0.1694	0.5749
Laws Feature 7	0.0037	0.0171	0.0067
Laws Feature 8	0.0008	0.0346	0.0151
Laws Feature 9	<0.0001	0.0753	0.0067
Laws Feature 10	0.0011	0.3924	0.0554
Laws Feature 11	0.2971	0.0058	0.0067
Laws Feature 12	<0.0001	0.3370	0.0215
Laws Feature 13	<0.0001	0.0952	0.0067
Laws Feature 14	0.2214	0.0033	0.0015

KEY RESEARCH ACCOMPLISHMENTS

1. A database of mammogram was generated containing 86 cases of serial mammograms, which were selected by reviewing 1334 cases. Based on this database, we further generated three datasets, i.e. missed cancer dataset, detected cancer dataset and normal dataset.
2. A series of statistical analyses of the computerized features of missed cancers (false negatives) versus detected ones (true positives) and their interval changes was taken. Based on the test P-values, the features with significant impact on radiologist's diagnosis and that potentially be useful for early detection could be identified.

REPORTABLE OUTCOMES

1. Presentation and/or proceedings paper

(a) Y. Qiu, L. Li, D. Goldgof, R.A. Clark, "Three dimensional deformation model for lesion correspondence in breast imaging," Proceedings of SPIE Medical Imaging, 2003.

2. Fundings Applied

(a) "Computer Aided Diagnosis of Focal Asymmetric Density", a project in Program Grant titled "Breast Imaging and Computerized Analysis Program" submitted to NCI, 2003.

CONCLUSIONS

This project is to explore an innovative CAD strategy for improving early detection of breast cancer in screening mammograms by focusing on computerized analysis and detection of cancers missed by radiologists. It is motivated by the facts that (1) it can be very instructive to review retrospectively the false negative results to determine why cancers were missed in mammographic screening; (2) some preliminary studies showed that there exist distinguishing features of missed cancer which is different from that of detected cancers. The research in first year is on data collection and analysis of characteristics of missed caner in terms of its computational features. By reviewing 1334 cases, a total of 86 missed cancer cases were collected which were used to generate three different datasets including mammograms with missed cancer, mammograms with screening-detected cancer and normal mammograms. A ground truth was generated by an experienced radiologist for feature extraction and analysis purpose. With the datasets and the ground truth, a variety of computerized features were extracted and analyzed to explore the difference of detected and missed cancer cases. A set of tests was applied to the extracted features individually from which the significant features distinguishing the missed cancer from detected ones could be identified and applied to the CAD design in next steps.

REFERENCES

[1] Yong Chu, Lihua Li, Dmitry Goldgof, Yan Qiu, Robert A. Clark, "Classification of masses on mammograms using support vector machine," Proc. of SPIE Medical Imaging, Feb. 2003.
[2] Stanton A. Glantz, *Primer of Biostatistics*, fifth edition, McGraw-Hill Medical Publishing Division, 2001.